

言葉とジェスチャーによる動きの制御方式

武藤 宏明 下岡 薫 松本 俊太郎 吉田 眞澄

筑波学院大学

経営情報学部 経営情報学科

まえがき

筆者らはこれまでに映像内の人物から関節単位のオプティカルフローを抽出し、それぞれの変化を歩容の概念で表現した。さらに、各関節の動きに対して13種類の言葉による速度と8方向の角度を結合させたコード化方式を開発し、その逆変換によって動きを復元した¹⁾。

本報告はこれらの技術を利用して、人間が与える指令情報によって、コンピュータ内の経路を適切に移動行動させる制御インタフェースに関する内容である。特に、人間がディスプレイ画面を見ながら言葉とジェスチャーで指示する知的な経路移動の制御を目的とした²⁾。

1. 基本的な考え

指令情報の策定では簡単な内容で明確な制御の実現を検討し、行動を構成する動きは可能な限り言葉、行動の開始と終了はジェスチャーで指示することにした。これにより行動の範囲を明確に規定できる。言葉は歩容にそって速度と方向で表現するが、ディスプレイ画面上では必ずしも適切に表現できない場合がある。これに対処するために歩容で表現できる動きは言葉、表現できない動きはジェスチャーで制御するようにした。その結果、不確定要素の大きい動きまで対応できると判断した。

2. 言葉によるインタフェース

言葉はディスプレイ画面を見ながら、音声で発声してコードに変換した。この際に、動きを13種類の速度と8種類の方向成分で入力するのではなく、両者を合成した言葉で発声させることでインタフェースの簡便化を図った。コード化に当たり、これまでに用いた言葉の表現と新たに採用した音声指令用の言葉を図1に示す。

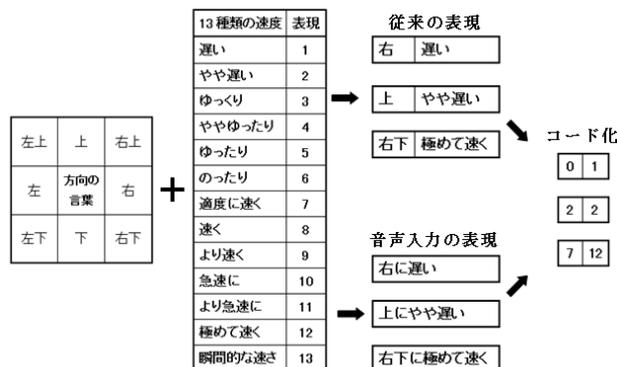


図1 従来の表現と音声入力の比較

3. ジェスチャーによるインタフェース

3.1 入力デバイス

ジェスチャーの認識は簡易で高認識率が期待できる KINECT を用いた。これにより関節単位の動きを表現してきたこれまでの情報表現と指示者による指令情報との親和性を達成させた。

3.2 開始と終了

開始と終了は行動全体を制御する命令であり、誰でもが納得して使用できるものでなければならない。そのジェスチャーとして、一定環境内の静止物体の誘導と、経路に沿って動物体を誘導する場合がある。筆者らは動物体の誘導に使用されている航空機のマーシャリングジェスチャーを採用することにした³⁾。そのジェスチャーを図2に示す。



(a) 開始の合図 (b) 終了の合図

図2 マーシャリングジェスチャー

Motion control scheme by words and gestures
TSUKUBA GAKUIN UNIVERSITY
The Faculty of Management & Information
The Department of Management & Information Studies

・ジェスチャー認識における特徴点

画面からの指示は着席して行うことを念頭において、KINECT で得られた関節単位の座標値に対し、肩より上部分で顔を除いた9箇所を特徴点とした。認識処理では関節部位名と各位置座標を対象とした。使用した特徴点を表1に示す。

表1 特徴点

		コード	
部位	部位	x座標	y座標
肩中央	sc	sc_x	sc_y
左肩	sl	sl_x	sl_y
左肘	el	el_x	el_y
左手首	wl	wl_x	wl_y
左手	hl	hl_x	hl_y
右肩	sr	sr_x	sr_y
右肘	er	er_x	er_y
右手首	wr	wr_x	wr_y
右手	hr	hr_x	hr_y

・識別の条件設定

両腕が“真っ直ぐ伸びている”，“交差している”が判断のポイントになるが、開始と終了以外の判断も重要である。そのために識別では各ジェスチャーの承認条件を盛り込んで、認識処理を実現した。

条件1：両腕の関節の相対的な位置

$$sl_y > el_y > wl_y > hl_y$$

$$\text{かつ, } sr_y > er_y > wr_y > hr_y$$

条件2：両腕の可動範囲

$$sc, sl(sr), el(er) \text{ のなす角: } 123^\circ \sim 90^\circ$$

$$sl(sr), el(er), wl(wr) \text{ のなす角: } 163^\circ \sim 180^\circ$$

$$el(er), wl(wr), hl(hr) \text{ のなす角: } 172^\circ \sim 180^\circ$$

角度は公開されている写真 50 枚を目視で整理し、統計処理した結果である。

条件3：交差の範囲

$$hl_x > hr_x$$

・判定論理

開始は条件1と条件2のANDかつ条件3のNOT、終了は条件1と条件3のANDとした。

3.3 歩容で表現できない動き

微妙な動きは局所的な部位として手を認識対象とした。KINECT では手のひらが関節と同等に扱われるので、その座標値のみを1秒間当たり30フレームで抽出することにした。

4. 評価実験

4.1 言葉による指令

歩容による指令情報の総数は104個の組み合わせになる。各情報に対して20回の発声実験を行い、正解率は99%であった。誤読は「速く」や「遅く」など文字列の短い言葉であった。参考実験として速度単独の言葉では98%であった。現行の音声認識技術では文字列が長いほど正解率が上がると判断する。

4.2 ジェスチャーによる指令

4.2.1 行動の開始と終了

開始と終了の各実験結果を図3に示す。いずれも正解の場合であり、図3中の点は抽出した特徴点、右側ほど識別が難しいジェスチャーである。

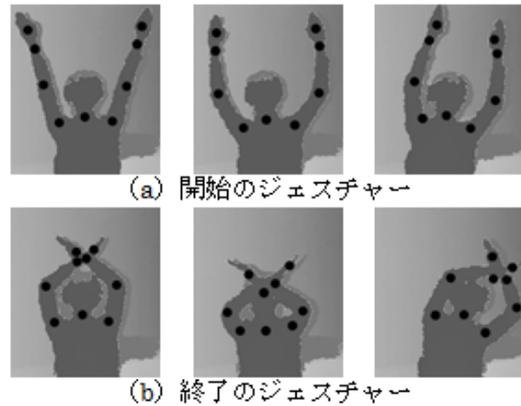


図3 開始と終了の認識結果

4.2.2 歩容で表現できない動き

微妙な動きを評価するために、あらかじめ曲線図形をディスプレイ画面に表示した。それに沿って手を移動させたときの軌跡との一致度合を検証することで精度を評価した。求めた座標点を制御命令としてコード化した速度と方向に変換した結果を表2に示す。曲線図形との一致度合は20回の試行において解像度640x480画素で、誤差は原図形に対してわずか±3%の範囲であった。

表2 算出した速度と方向

速度	9.9	7.8	8.5	7.6	11.0	11.7	3.6	13.0	10.8
方向	45.0	50.2	69.4	66.8	84.8	59.0	56.3	32.5	21.8

5. まとめ

行動を指示する情報として、言葉とジェスチャーの認識について考察した。言葉は歩容の速度と方向を合成した音声認識によってコード化し、ジェスチャーは行動の開始と終了、さらに言葉で表現できない部分を人間の手で制御できる見通しを得た。最後に、日頃ご指導いただく井高悠斗、渡辺悠太両氏に深謝する。

参考文献

- 1) 井高, 鈴木, 武藤, 吉田; 歩容の情報化による動きの表現, 情処学会全大76回6Q-6.
- 2) 下岡, 武藤, 井高, 吉田; 言葉とジェスチャーによる行動の制御システム, 情処学会全大77回3R-5.
- 3) Australian Government Department of Defence; <http://www.defence.gov.au/> (2014-12-28).